

Réseaux Bayésiens Dynamiques pour la reconnaissance des caractères imprimés dégradés

Khalid HALLOULI¹, Laurence LIKFORMAN-SULEM¹, Marc SIGELLE¹

¹ École Nationale Supérieure des Télécommunications
Département de Traitement du Signal et des Images
46 rue Barrault 75634 Paris Cedex 13 France

khalid.hallouli@enst.fr, likforman@enst.fr, marc.sigelle@enst.fr

Résumé – Le but de ce travail est de présenter une nouvelle approche pour la reconnaissance des caractères imprimés dégradés. Notre approche consiste à construire deux chaînes de Markov cachées [HMMs] à l'aide des réseaux bayésiens dynamiques, nommées HMM vertical et horizontal. Un HMM-vertical (respectivement HMM-horizontal) est un modèle qui prend pour séquence d'entrée les colonnes de pixels du caractère (respectivement les lignes de pixels). Nous couplons ensuite ces chaînes suivant deux modèles de couplage en utilisant les réseaux bayésiens dynamiques. Les résultats expérimentaux montrent que les modèles de couplage augmentent le taux de reconnaissance de 8 % à 10 % relativement au système de reconnaissance utilisant les modèles non couplés.

Abstract – This paper presents a new approach for printed character recognition. Our approach consists in building two Hidden Markov Models (HMMs) using Dynamic Bayesian Networks (DBN). The first HMM is constructed using a vertical flow of writing (vertical HMM), the second one is obtained using a horizontal flow (horizontal HMM). And then we couple these HMMs according to two models of coupling. The experiments show that the coupled HMMs models increase the recognition rate from 8% to 10% relative to the system of recognition using separate HMMs models.

1 Introduction

Les modèles probabilistes, tels les modèles de Markov cachés (HMMs), sont largement utilisés pour la modélisation et la reconnaissance de l'écriture, car ils savent bien s'adapter à la variabilité des observations. Nous appelons un HMM vertical (respectivement horizontal) un modèle qui prend pour séquence d'entrée les colonnes de pixels du caractère (respectivement les lignes de pixels). Ces modèles donnent des performances acceptables et les HMMs verticaux sont plus performants que les HMMs horizontaux [2]. Une approche basée sur la fusion de décision de ces deux HMMs est très utile car elle regroupe l'information sur les colonnes et sur les lignes, ce qui n'est pas le cas d'un HMM simple. Cependant cette approche est loin d'être optimale. En effet les deux HMMs vertical et horizontal sont supposés indépendants, ce qui semble peu réaliste. De plus, l'étape de recombinaison des résultats des deux classificateurs HMMs est à optimiser dans le cas de la reconnaissance. Le but de cet article est de proposer une nouvelle approche pour la reconnaissance des écritures (caractères imprimés dégradés). Dans notre approche nous allons modéliser les dépendances entre les nœuds (variables aléatoires d'états) en créant des interactions entre les deux HMMs. Pour cela nous utilisons le formalisme des Réseaux Bayésiens (Bayesian Networks BNs). Ce dernier est devenu un formalisme puissant qui unifie différents concepts de modélisation probabiliste utilisés en statistique, intelligence artificielle, reconnaissance de la parole[1,2,4,6]. Il est couramment employé pour le raisonnement sous incertain et peut être considéré comme un système expert probabiliste. Les BNs constituent un cadre intéressant pour deux raisons. D'une part grâce à leur structure graphique les BNs offrent un

outil naturel pour représenter les dépendances entre les différentes variables d'un système donné. D'autre part, en exploitant les indépendances conditionnelles entre les variables, ils introduisent une certaine modularité dans les systèmes complexes. Ainsi les BNs fournissent un outil séduisant pour modéliser des systèmes complexes et conduisent à des algorithmes rapides d'inférence et d'apprentissage.

Nous allons présenter deux nouveaux modèles de couplage où nous couplons les deux HMMs vertical et horizontal, ensuite nous développons des algorithmes d'apprentissage et de décodage pour ces deux nouveaux modèles, et nous conduisons des expériences pour illustrer leur potentiel.

2 Réseaux Bayésiens

Un Réseau Bayésien (réseau probabiliste ou Bayesian Belief Network, ...) [4, 5, 13] est un modèle représentant des connaissances incertaines sur un phénomène complexe, et permettant à partir des données un véritable raisonnement. Ainsi un BN est un graphe acyclique orienté dont les nœuds sont des variables aléatoires qui ont des valeurs discrètes ou continues selon une loi Gaussienne. Les deux notions essentielles que modélisent les Réseaux Bayésiens sont l'incertitude et la causalité. Formellement un BN statique est défini par la donnée de deux éléments : un graphe acyclique G et une paramétrisation θ .

Soient un ensemble de variables aléatoires $X = (X_1, X_2, \dots, X_N)$ et $P(X)$ sa distribution jointe de probabilité. Le graphe G indique les indépendances conditionnelles entre les nœuds (variables aléatoires). La paramétrisation θ est donnée en terme de probabilités conditionnelles des nœuds con-

naissant leurs parents. Donc la probabilité jointe peut être écrite sous la forme

$$P(X_1, X_2, X_3, \dots, X_N) = \prod_{1 \leq i \leq N} P(X_i | C(X_i)) \quad (1)$$

où $C(X_i)$ est l'ensemble des causes (parents) de X_i dans le graphe G . La sémantique des indépendances conditionnelles d'un BN implique qu'une variable est indépendante de toutes les autres variables du réseau en connaissant ses parents, à l'exception de ses descendants.

Les Réseaux Bayésiens Dynamiques (DBN) [17] sont une extension des réseaux Bayésiens qui représente l'évolution temporelle ou spatiale des variables aléatoires. On considère un ensemble $X_t = (X_t^1, \dots, X_t^N)$ de variables évoluant dans $[0, T]$. La probabilité jointe peut être codée par un réseau statique avec $(T + 1) * N$ variables avec une possibilité d'avoir une structure ou des paramètres différents à chaque instant t . Si le processus est stationnaire, les hypothèses d'indépendance et les probabilités conditionnelles associées sont identiques pour tous les temps t . Dans ce cas un DBN peut être représenté par un BN dont la structure est répétée à chaque instant t . Ainsi un HMM est un cas particulier de DBN (voir Fig.1)

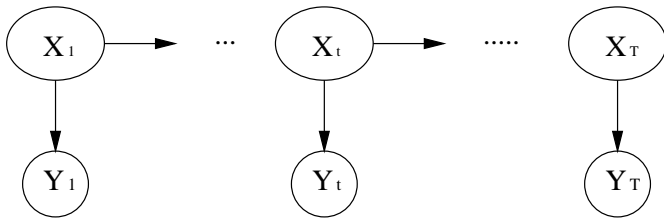


FIG. 1 – HMM représenté comme un DBN avec les $(X_t)_{1 \leq t \leq T}$ représentant les états cachés et les $(Y_t)_{1 \leq t \leq T}$ représentant les observations

3 Modélisation



FIG. 2 – exemples de caractères

Étant donné un caractère (exemple Fig.2), l'idée principale est, qu'au lieu de considérer un HMM indépendant vertical ou horizontal, on construit un DBN plus complexe (uniforme) en couplant deux HMMs. Un HMM-vertical (respectivement HMM-horizontal) est un modèle qui prend pour séquence d'entrée les colonnes de pixels du caractère (respectivement les lignes de pixels). Ce couplage est réalisé en ajoutant des liens (orientés) entre les variables (nœuds). Un des problèmes rencontrés est : quels sont les liens à ajouter (autrement dit, quelle est la structure du réseau)? La meilleure solution est d'ap-

prendre la structure graphique à partir des données. Cette stratégie, appelée *apprentissage structurel* [12], n'est pas l'objectif de notre approche. Notre objectif est de fixer une structure graphique puis d'évaluer si elle est prometteuse.

3.1 Structure graphique

Nous fixons deux structures en se basant sur les critères suivants :

- le modèle construit doit avoir un nombre raisonnable de paramètres pour que la complexité des calculs reste abordable.
- aucune variable continue ne doit avoir de fils discret afin de pouvoir appliquer un algorithme d'inférence exacte (JLO)[13, 16].
- des liens doivent exister entre les variables cachées.

On note, pour $j = 1, 2$ $(X_t^j)_{1 \leq t \leq T}$ les variables d'états cachées et $(Y_t^j)_{1 \leq t \leq T}$ (respectivement $(\bar{Y}_t^j)_{1 \leq t \leq T}$) les séquences d'observations correspondant aux colonnes des pixels du caractère (respectivement lignes) comme le montre la figure Fig.3.

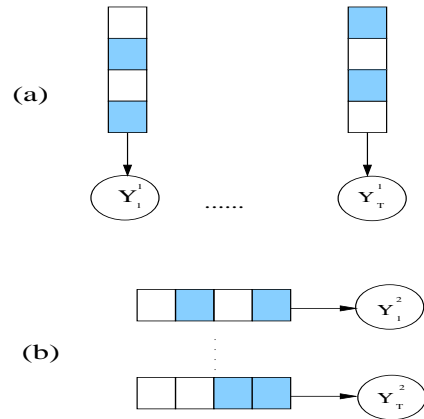


FIG. 3 – Séquence d'observations : (a) selon les colonnes et (b) selon les lignes

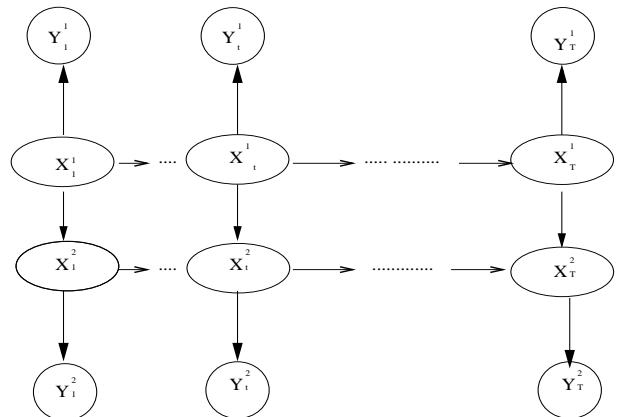


FIG. 4 – premier modèle de couplage(DBN) : DBN1

Les deux figures Fig.4 et Fig.5 présentent deux modèles de couplage, DBN1 et DBN2 qui satisfont aux critères annoncés . Les $X_t^j = (X_t^j)_{1 \leq t \leq T}$ pour $j = 1, 2$ sont des variables discrètes prenant leurs valeurs dans un ensemble $I = 1, \dots, k$

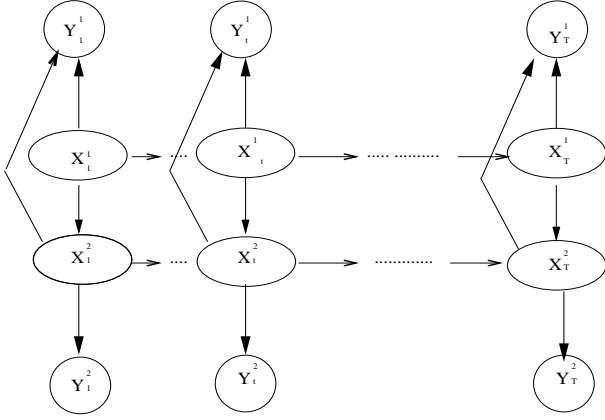


FIG. 5 – deuxième modèle de couplage(DBN) : DBN2

avec k le nombre des états cachés, Y_t^1 (respectivement Y_t^2) est un vecteur aléatoire continu suivant une loi gaussienne représentant l'observation des fichiers des caractères correspondant aux colonnes de pixels (respectivement aux lignes de pixels) à chaque position t .

- **DBN1** : dans ce DBN les variables d'états cachées $(X_t^1)_{1 \leq t \leq T}$ sont liées aux variables d'états $(X_t^2)_{1 \leq t \leq T}$. Pour chaque $j \in \{1, 2\}$ les états (X_t^j) agissent sur les observations (Y_t^j) . Pour chaque caractère et pour $(i, j, k) \in I^3$, la paramétrisation numérique θ est définie par la donnée de A^1, U^1, b^1 et b^2 :

$$\begin{cases} A_{i,j}^1 = P(X_t^1 | X_{t-1}^1) \\ U_{i,j,k}^1 = P(X_t^2 = k | X_{t-1}^2 = i, X_t^1 = j) \\ b_i^l(\cdot) = P(Y_t^l = \cdot | X_t^l = i) \text{ pour } l = 1, 2 \end{cases} \quad (2)$$

- **DBN2** : dans ce DBN il y a un lien supplémentaire par rapport au premier modèle c.a.d les états (X_t^2) contrôlent les deux observations lignes et colonnes (Y_t^1) et (Y_t^2) . Pour chaque caractère et pour $(i, j, k) \in I^3$ La paramétrisation numérique θ est définie par la donnée de A^2, U^2, b^1 et b^2 :

$$\begin{cases} A_{i,j}^2 = P(X_t^1 | X_{t-1}^1) \\ U_{i,j,k}^2 = P(X_t^2 = k | X_{t-1}^2 = i, X_t^1 = j) \\ b_{i,j}^1(\cdot) = P(Y_t^1 = \cdot | X_t^1 = i, X_t^2 = j) \\ b_i^2(\cdot) = P(Y_t^2 = \cdot | X_t^2 = i) \end{cases} \quad (3)$$

Contrairement à un HMM (vertical où horizontal), ces deux modèles sont des modèles bi-dimensionnels qui tiennent compte de l'information selon les lignes et les colonnes.

3.2 Estimation des paramètres

Dans les expériences que nous effectuons, l'apprentissage des modèles est effectué de façon indépendante modèle par modèle, en utilisant l'algorithme EM (Expectation Maximisation). L'étape E de l'EM est réalisée en utilisant l'algorithme d'inférence JLO [2]. Les formules de re-estimation des paramètres peuvent être trouvées dans [9].

4 Résultats expérimentaux

Les expériences sont effectuées sur 2080 fichiers de caractères extraits de la base de données UW English Document Image Database [7]. Les images des caractères extraits de cette

base, sont normalisés à 50x50 pixels. Toutes les expériences sont faites à l'aide de la Toolbox en Matlab nommée : Bayesnet [8], et ces expériences sont réalisées dans le cas dynamique avec 26 classes (modèles). Les premières expériences consistent à trouver le nombre d'états optimal pour les HMMs verticaux et horizontaux et les deux modèles de couplage DBN1 et DBN2 (voir Fig.6). Ainsi le résultat final concernant les taux de re-

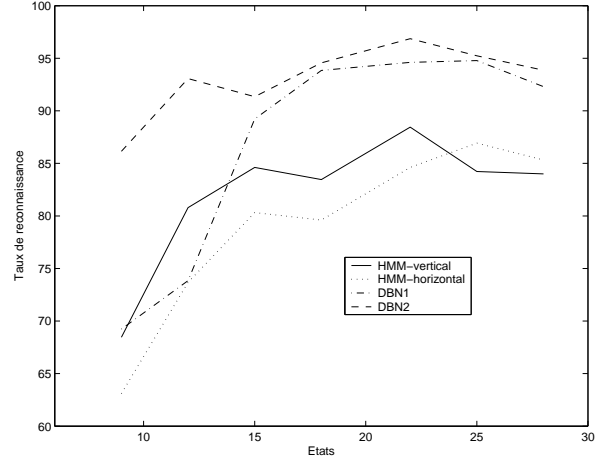


FIG. 6 – Évolution de taux de reconnaissance par rapport au nombre d'états

connaissance, est donné dans le tableau Tab.1. On constate que

TAB. 1 – Résultats de reconnaissance pour les différents modèles

| | Taux de reconnaissance |
|----------------|------------------------|
| HMM-vertical | 88.56 |
| HMM-horizontal | 85.56 |
| DBN1 | 94.76 |
| DBN2 | 96.84 |

les deux modèles de couplage sont meilleurs par rapport aux HMMs simples (voir Tab.1). Cela est dû fait que les HMMs simples ne prennent en considération que l'information suivant la ligne ou bien suivant la colonne, tandis que les modèles couplés prennent en compte l'aspect 2D de l'image. Les deux modèles DBN1 et DBN2 accordent plus d'importance aux observations de type colonne. Cette dissymétrie se retrouve au niveau des performances des systèmes HMM-vertical et HMM-horizontal. Le modèle DBN2 accorde encore plus d'importance aux colonnes de pixels car la variable d'état des lignes est influencée par l'observation des colonnes, ce qui explique la différence entre les taux de reconnaissance des modèles couplés. Les expériences montrent que sur 1040 images de caractères, les modèles de couplages DBN1 et DBN2 augmentent le taux de reconnaissance de 8% à 10% relativement au système de reconnaissance utilisant les modèles non couplés.

5 Conclusion

Dans cet article nous comparons expérimentalement les deux modèles de couplage : DBN1 et DBN2 avec les deux HMMs

vertical et horizontal en se basant sur la vraisemblance. Les résultats obtenus sont très encourageants et beaucoup de détails demandent encore à être améliorés, surtout en ce qui concerne les conditions initiales, les algorithmes d'inférence et la structure optimale. Pour ce dernier point des travaux sont en cours pour trouver d'autres stratégies structurelles (graphes) qui pourraient améliorer les performances obtenues.

Références

- [1] S.L. Lauritzen, D.J. Spiegelhalter. *Local computations with probabilities on graphical structures and their application to expert systems* J.Roy.Stat.Soc. B, 50,157-224,1988.
- [2] F.V. Jensen,S.L. Lauritzen,K.G. Olsen *Bayesian updating in causal probabilistic networks by local computations* Comp.Stat.Quart., 4, 269-282, 1990.
- [3] K. Hallouli, L. Likforman-Sulem, M. Sigelle. *A comparative study between decision fusion and fusion data in Markovian printed character recognition* In 16th International Conference on Pattern Recognition Quebec, Canada , August 2002.
- [4] J. Pearl. *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference* Second edition, Morgan Kaufmann, San Mateo, Calif, 1988.
- [5] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter. *Probabilistic Networks and Expert Systems* Springer-Verlag New York, 1999.
- [6] A. Becker, P. Naïm. *Les réseaux bayésiens* Eyrolles, Paris , 1999.
- [7] English DOCUMENT database I and II ,CD ROM set, 1993
- [8] K. Murphy. *Bayesnet tutorial*,
<http://www.ai.mit.edu/~murphyk/Bayes/bnintro.html>, October 2001.
- [9] K. Daoudi, D. Fohr, C. Antoine. *Continuous Multiband Speech Recognition using Bayesian Networks* IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Trento Italy, December 2001.
- [10] K. Daoudi, D.Fohr, C. Antoine. *A new approach for multiband speech recognition based on probabilistic graphical models* In International Conference on Spoken Language Processing (ICSLP), Beijing, China, October 2000.
- [11] K. Daoudi, D. Fohr, C. Antoine. *Réseaux Bayésiens pour la Reconnaissance Multi-bandes de la Parole* in XXIV ème Journées d'Études sur la Parole, Nancy Juin 2002.
- [12] M. Deviren, K. Daoudi. *Structural learning of dynamic Bayesian Networks in Speech Recognition* In EUROSPEECH , Alborg, Denmark, September 2001.
- [13] K. Murphy. *Dynamic Bayesian Representation, Inference and Learning* PhD thesis , University of California, Berkeley 2002.
- [14] M.I. Jordan, T.J. Sejnowski. *Graphical Models : Foundations of Neural Computation* MIT Press, Massachusetts Institute of Technology 2001.
- [15] C. Choisy, A. Belaid. *Couplage d'une vision locale par HMM et globale par RN pour la reconnaissance de mots manuscrits* pages 345-354, CIFED, 2002.
- [16] F.V. Jensen. *Bayesian Networks and Decision Graphs* Springer-Verlag New York, 2001.
- [17] T. Dean, A. Kanazawa. *A model for reasoning about persistence and causation* Artificial Intelligence, 93(1-2) : 1-27, 1989.
- [18] A. Elms, S. Procter, J. Illingworth. *The advantage of using an HMM based approach for faxed word recognition* in IJDAR'98, volume 1, pages 18–36, 1998.